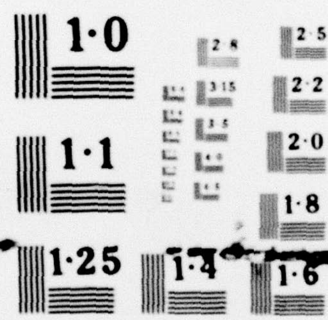
AD-A075 594 GEORGE WASHINGTON UNIV ALEXANDRIA VA HUMAN RESOURCES--ETC F/G 12/1
A METHOD FOR COMPUTING THE KENDALL TAU COEFFICIENT,(U)
NOV 53 H F BRIGHT

UNCLASSIFIED NL

1 OF 1
AD-
A075594

END
DATE
FILMED
11-79
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

[redacted]

(12) 9

(6)

NR

A Method for Computing the Kendall Tau Coefficient

by

(10) Harold F. Bright

Published in: Educ. and Psych. Measurement, 1954, Vol. 14, page 702.

Human Resources Research Office
Washington, D. C.
(11) 24 Nov[redacted]53

15

173 200

DEPARTMENT OF THE ARMY
ARI FIELD UNIT, BENNING
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
P.O. BOX 2086, FORT BENNING, GEORGIA 31905

PERI-IJ

8 August 1979

SUBJECT: Shipment of Documents

Defense Documentation Center
Cameron Station
Alexandria, VA 22314
ATTN: Selection & Cataloging

The Documents in these shipments are approved for public release. The distribution is unlimited.

FOR THE CHIEF:

Alexander Nicolini

ALEXANDER NICOLINI
Major, Infantry
R&D Coordinator

It sometimes happens that correlational analysis can best be handled by rank methods since the variate values are expressed as ranks. In other cases, rank methods may be preferred by the investigator because of peculiarities in the bivariate distributions involved. In such cases the Spearman Rho Coefficient has been used quite commonly despite some difficulties with tests of significance. The Kendall Tau Coefficient possesses a frequency distribution which is easy to calculate for small n and which approaches normality quite rapidly.[1] However, the computation of the tau coefficient from its definition is tedious. A method which can be carried out easily and quickly for small numbers of cases, say fifty or less is, therefore, set forth in this paper.

The tau coefficient is quite simply defined. Given ranks assigned to a set of individuals on two tests as follows:

Individuals	1	2	3	4	5	6
Test A	1	3	5	4	2	6
Test B	2	1	3	5	4	6

Consider each possible pair of ranks in the A array. Assign a plus one to every pair which is in normal order and a minus one to each pair which is in inverted order. Do the same for the B array. Then for corresponding pairs of the two arrays, the plus or minus one values are multiplied together and the products are summed. The result, in this case 7, is denoted by S. The total number of pairs in each array must be the com-

----

[1] Kendall, Maurice G. Rank Correlation Methods. (London: Charles Griffin and Company, Limited, 1948.) pp.37ff.

binations of the n ranks taken two at a time, in this case $\binom{6}{2} = 15$. Then the tau coefficient is defined as

$$\tau = \frac{S}{\binom{n}{2}} = \frac{7}{15} = .467$$

Kendall points out that an easier procedure is to rank one array in normal order so that the two arrays previously given would appear as follows:

A:	1	2	3	4	5	6
B:	2	4	1	5	3	6.

Then, since array A can contribute only positive pairs to the value of S, it is only necessary to consider the contributions of the pairs in array B. It is clear that $S = P - Q$ and $\binom{n}{2} = P + Q$ where P is the number of positive pairs in B and Q is the number of negative pairs. A simple algebraic transformation then gives the result

$$\tau = \frac{S}{\binom{n}{2}} = 1 - \frac{2Q}{\binom{n}{2}} = \frac{2P}{\binom{n}{2}} - 1.$$

It is evident that the Q score can be arrived at by counting the number of transpositions necessary to put array B into normal order. In the present example the array is in the order

2 4 1 5 3 6.

It takes two transpositions to put the 1 in its proper initial position so that the array becomes

1 2 4 5 3 6.

The 2 is already in normal position, and two transpositions bring the 3 into proper position to give

1 2 3 4 5 6.

Now the array is in normal order having required four transpositions to transform it. Thus $Q = 4$ and $\tau = .467$

the same result as was arrived at by the previous method.

This is an adequate method except for the disadvantage that the writing down of the new array each time is quite laborious, particularly if a larger number of rankings is involved. A simple device for accomplishing the computation can be constructed in a few minutes.

A trough is constructed of depth and width convenient to accomodate a set of numbered blocks. For convenience the drawing of Figure 1 shows only six blocks. The procedure can be used with reasonable ease for as many as fifty cases.

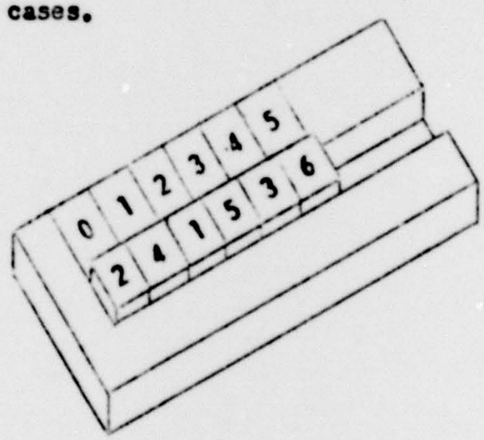


FIGURE 1

The trough is set on a slant so that as one of the blocks to the left is taken out the others will slide to the left. The upper edge of the trough is marked off into divisions equal in width to the blocks and numbered from the left starting with the number 0. All the rankings pertaining to each individual studied are entered upon a card. It is an easy matter to arrange the cards in order according to one of the variables ranked. For instance, the rankings previously mentioned could be used in which the cards would be placed in order for the A array and the B array would then appear as

2 4 1 5 3 6.

Then the blocks are placed in the trough in the order of the B array. The numbers on the side of the trough then correspond to the A array numbers one position out of phase.

The next step is to remove the 1 block from the trough, at the same time observing the number of the space on the side of the trough corresponding to it which in this case is 2. The value 2 is entered in an adding machine as part of the value of Q since it represents the number of transpositions necessary to place the 1 block in its normal position. The block numbered 2 is then removed and placed next to the 1 block on the table. It needs no transpositions to place it in its normal position as is shown by the zero opposite it. The 3 block is then removed from its new position opposite the 2 on the side of the trough and the 2 value is entered in the adding machine to represent the number of transpositions needed to place the 3 block in its normal position. The 4, 5 and 6 blocks are then removed consecutively from the zero position and since no transpositions are required for them, nothing is entered in the adding machine. The total in the machine then gives the total Q score. In a simple case such as this one the adding machine is not really needed but when the process is applied to twenty or thirty blocks it is much easier to use an adding machine than to carry the addition forward mentally.

The above process is an easy, almost fool-proof method of computing the Q score. It is executed much more easily when two operators work. Operator X calls the ranks for variable B off the cards arranged in order of variable A. Operator Y places the blocks in the trough in the proper order. Then Operator Y takes the blocks out in order of rank calling out the position value from which each block is taken so that Operator X can enter it in the adding machine. As each block is removed all blocks on its right move to the left. With a manually operated adding machine it is easy to compute Q scores quite rapidly for as many as fifty ranks. When the blocks are all out they are then in order on the table and it is an easy matter to repeat the process for variable C, D, etc. When all the Q scores have been computed for the possible combinations of variable A with other variables, the cards are sorted again in order for variable B and the process is repeated to include all the combinations of B with all variables except A. This procedure is repeated until all possible Q scores have been computed. It is then an easy matter to compute the tau coefficients.

Slight modifications of the procedure are necessary in the case of tied ranks. Two different situations may be present. In the first, one variable has no tied ranks. In such cases it is most practical to sort the cards on the untied ranking. Then suppose that a tie occurs which uses up ranks 4, 5 and 6. The rank 5 is assigned to all three. But in the present method, it is only necessary to use the three blocks numbered 4, 5 and 6, making sure that they appear in order in the trough from left to right. This makes certain that as they are taken out of the trough none of them will affect the Q score contributions of the others. When several ties appear in the ranking it is usually useful to keep the ranking in front of the operator so that he will be sure to get the blocks in the correct order when he places them in the trough.

In some cases ties may appear in the variable for which the cards are sorted. In such a case it is necessary to look through the cards before arranging the blocks and to make sure that the ties on this variable are arranged in order of size of the second variable. This gives a unique answer in every case so that if this procedure is followed one arrives at the same result as he would if the cards were sorted on the untied variable. This procedure is always necessary, of course, if ties appear in both variables. It is quite obvious that the easiest method for, say, twenty variables is to sort the cards first on the variables which have no ties so that at the end only a few need to be handled in the more cumbersome manner.

The denominator of tau also must be adjusted in case the ranking of either variable contains ties. The denominator will then be not the total number of combinations in pairs of the rankings but will be this value less the number of tied pairs since tied pairs contribute nothing to S. This matter is treated fully by Kendall.[2]

---

[2] *op. cit.* Chapter 3.

A METHOD FOR COMPUTING THE KENDALL
TAU COEFFICIENT

By Harold F. Bright

Name Date